Journal of Agricultural Studies
ISSN 2166-0379
2015, Vol. 3, No. 1

Application of Spectral Methods for Synthesis of Symptoms for Qualification of Agricultural Products

Radoslava Nikolova Gabrova (Corresponding author)

Dept. of Computer systems and technologies, University of Food Technologies

bul. Maritza 26, Plovdiv, Bulgaria

E-mail: rgabrova@yahoo.com

Lena Filipova Kostadinova

Dept. of Computer systems and technologies, University of Food Technologies

bul. Maritza 26, Plovdiv, Bulgaria

E-mail: lenakostadinova@yahoo.com

Received: August 11, 2014 Accepted: August 26, 2014

doi:10.5296/jas.v3i1.6117 URL: http://dx.doi.org/10.5296/jas.v3i1.6117

Abstract

The basic stages during automatic qualification of agricultural products are: acquisition of initial information of the quality state, mathematical proceeding of this information and decision making for object qualification to defined sets according to the state of quality. One of the steps in the second stage is transformation the patterns of qualified products to a new space of symptoms. In this paper an approach for transformation the initial description of the objects to a new space of spectral symptoms, independently of the size of the initial description of the object is presented. A relation for synthesis of symptoms for recognition has been generated.

Keywords: Qualification of agricultural products, Spectral methods, Synthesis of symptoms

1. Introduction

In the foundation of the process automatic qualification and sorting of agricultural products is the task for recognition (classifying) some unknown state (of the object under classification) from several known states of the reference objects presenting in advance defined classes of quality. The reference objects form so called “training set” necessary for the synthesis of any classifier.

The initial information of the quality state of agricultural products most often is obtained by using opto-electronic methods. The reflected or transmitted through the object light, forms image of the product under qualification as distribution of brightness, which transforms to electrical signal. By the reason of using up to date digital systems for implementation the algorithms and programs for classification it is necessary as a next stage of proceedings to realize discretization in time and quantization levels of the initial signals.

As a result the descriptions of the objects are presented as numerical sequences. At one-dimensional case each object is presented by a vector with coordinates measured values of a signal at reporting points i:

$$\mathbf{s}(\mathrm{i}) = \{\mathrm{s}(0),\ \ \mathrm{s}(2),\ \ \mathrm{s}(3),\ \ \ldots\ \ \mathrm{s}(\mathrm{i}),\ \ \ldots,\ \ \mathrm{s}(\mathrm{N}-1)\}^{\mathrm{T}}. \tag{1}$$

Each of the reported values contains information about the quality state of the certain part of the product volume. In this sense the particular value s(i) can be considered as a symptom of quality in this region of the product. The general state of quality will be determined as an integrated assessment with contribution of all N symptoms.

The more symptoms the more precisely they will present the characteristics of the object which are important for the process of quality recognition. At the same time the dimension of the symptom space is connected with the data volume and complexity of the computing procedures that affect the performance and the possibility for on-line quality recognition. Therefore there is a conflict concerning the greater number of symptoms for the accurate determination the quality state of the objects, and the smaller number of symptoms for implementation in practice the quality recognition on-line.

The avoidance of the conflict can be achieved by transformation the initial symptom description (space) to a new one with less number of symptoms. This process is known in the theory of pattern recognition as “minimization of the initial description of the objects”, “rationalization of the symptoms space” or “compression the data”. The main requirements for the new space are to have simplified formalized descriptions of the symptoms and high value of contained information concerning quality state of the objects.

Universal approach for the choice of new symptoms space is not known. One of the methods is a heuristic synthesis of the functions for transformation. Using this approach the end result will depend on the knowledge, the experience and the intuition of the experts, working on the problem. The other feature is a fact that the heuristic synthesised symptoms are correlated and it is difficult to make a selection on the base of their information value. The other approach that avoids the disadvantages of the heuristic is based on strictly mathematical (theoretical) methods for forming the symptoms space. As a direction in this approach is the development of methods for symptom synthesis based on the spectral methods of functional analysis. in this method signals discrete time functions are transformed to new space possessing own precise defined coordinates (coordinate Basis). This coordinate Basis is described by a system of linearly independent functions (basis functions) acting as a transformant (operator) of conversion.

The objects under testing (agricultural products) have different geometric characteristics and the initial information of the quality state obtained during their longitudinal scan will be as one-dimensional or two-dimensional electric signals representing random functions of time with different duration. This circumstance is embarrassing for the application of the approach for synthesis of symptoms by using the orthogonal transformations in qualifying and sorting agricultural products. The work proposes an approach for elimination this embarrassing circumstance.

2. Exposition

For the synthesis of a new complex of symptoms for recognition can be used orthogonal transformations (from the functional analysis) on the initial description of the tested objects (Georgiev). After the transformations the initial information is presented as a sum of scaled (weighted) orthogonal functions with coefficients obtained from the decomposition. This allows as symptoms for recognition to be used namely the coefficients from decomposition. One of the advantages of this transformation is the fact that a decomposition to sum of scaled orthogonal functions is used and that automatically eliminates correlation of the symptoms – the spectral constituents and there is no need to test their statistical connectivity. Also using this approach there is no need to test the informativity of symptoms insofar the spectral analysis a priory has given a decision of the problem – the spectral constituents with greater numbers carry a small part of the energy of the signal and in this sense they are low informative relative to the image.

Most frequently are used Fourier, Walsh, Hadamard, Hartley, Haar, cosine and other transforms. These transformations applied on the initial description (1) of the tested object can be written as (Donevska):

$$S(n) = \sum_{i=0}^{N-1} s(i)a(i,n), \quad n = 0,1,\dots,N-1 \tag{2}$$

$$s(i) = \sum_{n=0}^{N-1} S(n)b(i,n), \quad i = 0,1,\dots,N-1. \tag{3}$$

The equation (2) presents the Fourier Transform, equation (3) – the Inverse Fourier Transform, a(i,n) is a core of the Fourier Transform, b(i,n) – core of the Inverse Fourier Transform and S(n) are coefficients of transformation.

In Discrete-Time Fourier Transform (DFT) the basic functions a(i,n) and b(i,n) are exponential and have respectively the type $\exp(-j2\pi in/N)$ and $[\exp(j2\pi in/N)]/N$ or for the DFT can be written (Donevska):

$$S(n) = \sum_{i=0}^{N-1} s(i)W_N^{ni}, \quad n = 0,1,\dots,N-1, \tag{4}$$

where $W_N^{ni} = e^{-j2\pi ni/N}$ is a core of the transformation (basic complex exponential functions), s(i) is the description of the object in the time domain and S(n) – in the frequency domain.

One of the properties of DFT when performed over a sequence of real numbers is that the zero coefficient (S(0)) and the coefficient S(N/2), if N is even number, are single in the output sequence, while all other coefficients S(i) are complex conjugated i.e. the amplitude spectrum of the signal is obtained symmetric. The algorithm itself offers reduction of symptoms, because as such for recognition are used modules (amplitudes) of the obtained from the decomposition complex exponential functions. The number of symptoms reduces to N/2+1 for N even number and to (N+1)/2 for N odd number. Subsequent reduction can be done by using the information value of the symptoms as a selection criterion.

For obtaining full information of the quality state of the tested objects, they are scanned throughout their volume (length). Taking into account the fact that agricultural products are complex biological objects, it can be concluded that formed descriptions are random sequences with different numbers of elements. These descriptions can be unified in size with adding zero reported values for the shorter objects. It is known from the theory of Fourier spectral analysis that adding zeroes to the time sequence do not change the representation in the frequency domain. In this case only the resolution changes – more points in the range from 0 to 2π. This approach (complementarity with zero reports) can not be used when using the coefficients as symptoms for recognition.

For example in accordance with (4), when n=0 the coefficient S(0) is determined by:

$$S(0) = s(0)W_N^0 + s(1)W_N^0 + ... + s(N-1)W_N^0 . \tag{5}$$

When unifying the size of the descriptions to N by adding zero elements, the value of the coefficient of decomposition S(0) will depend on the number of actual reported values s(i) (without zero elements) and will substantially differ from the products of the same class. These features require an adaptation of the method for symptom synthesis based on orthogonal basic Fourier functions when the initial descriptions of objects have different number of coordinates. It is proposed the transformation of the initial description to be performed according to the number of actual reported values.

When the descriptions of objects have different number of reported values in the sum (5) participate different number of addends and therefore the coefficients of decomposition S(0) can not be compared directly, but have to be normalized with respect to the number of points in their description and to use as one of the symptoms for recognition the coefficient $S^*(0)$:

$$S^*(0) = \frac{1}{N}\sum_{i=0}^{N-1} s(i)W_N^{0.i} . \tag{6}$$

Thus the coefficient $S^*(0)$ brings a sense of the mean value. Such a physical meaning has the most widely used heuristic symptom.

In accordance with (4) at n=1 the coefficient S(1) will be determined:

$$S(1) = s(0)W_N^0 + s(1)W_N^1 + s(2)W_N^2 + \dots + \\ + s(N/2)W_N^{N/2} + S(N/2+1)W_N^{N/2+1} + \dots + S(N-1)W_N^{N-1} . \quad (7)$$

Using the properties of rotation multiplier W_N^{nk} (Donevska):

$$W_N^{kN} = 1, \quad W_N^{p+kN} = W_N^p, \quad W_N^{N/2} = -1, \quad W_N^{p+N/2} = -W_N^p, \quad (8)$$

If N is even number from (7) the coefficient S(1) is obtained:

$$S(1) = (s(0) - s(N/2))W_N^0 + (s(1) - s(N/2+1))W_N^1 + \\ + (s(2) - s(N/2+2))W_N^2 + \dots + (s(N/2-1) - s(N-1))W_N^{N/2-1} . \quad (9)$$

One typical image of an agricultural product, containing 16 reported values is shown in Figure 1. The eight differences analytically described in the equation (9) are designated in the fig.1 by numbers from 0 to 7. The differences with numbers from 0 to 3 will have negative values (presented in the Figure 1 in red color) and the differences with numbers from 4 to 7 have positive values (presented in the fig.1 in green color). Taking into account formula (9) and the shape of the image in fig.1 the coefficient S(1) can be considered as a symptom of quality assessing symmetry of the description that will be impaired by the presence of defects.

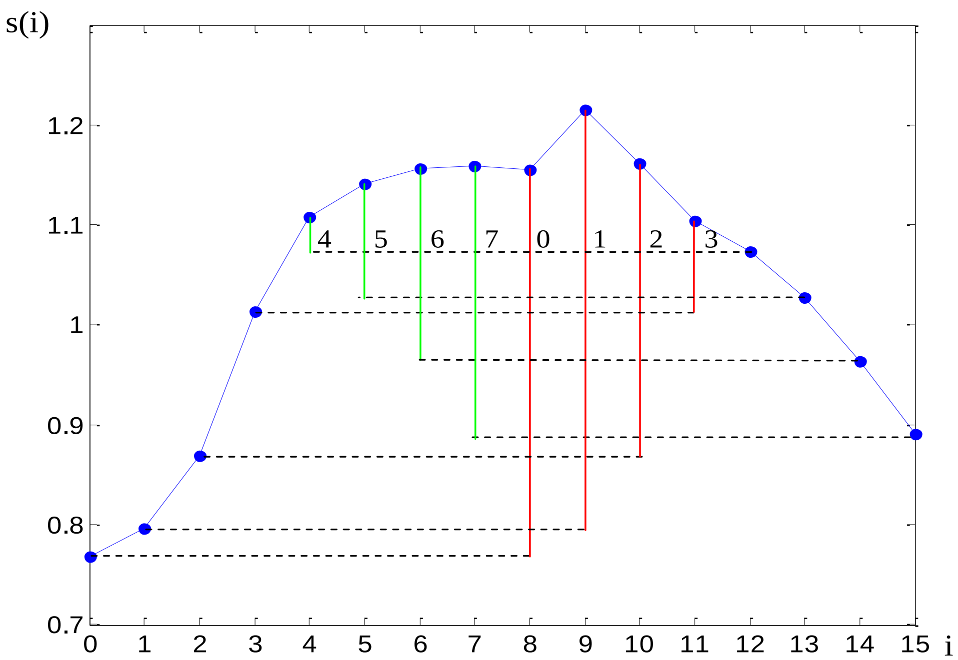

Fig. 1 Typical image of an agricultural product with 16 reported values in the description and differences from 0 to 7 according to equation (9)

What happens if the image of the object contains 12 reported values and the description is supplemented with 4 zero elements so that all become 16 when using DFT? The differences from 4 to 7 will not have the same signification as in the case with 16 reported values and the coefficient S(1) can not be considered as a symptom assessing symmetry. Again it can be concluded that DFT should be done with the description containing the actual reported values (without supplementation with zero elements). In order to have comparable results the coefficient S(1) should be normalized with respect to the number of actual reported values

$$S^*(1) = \frac{1}{N}\sum_{i=0}^{N-1} s(i)W_N^{1.i} \,. \qquad (10)$$

According to equation (4) when N even number after application of the properties (8), for the coefficient S(N/2) can be written:

$$\begin{aligned} S(N/2) = s(0)W_N^0 - s(1)W_N^0 + s(2)W_N^0 - s(3)W_N^0 + \ldots + \\ - s(N-3)W_N^0 + s(N-2)W_N^0 - s(N-1)W_N^0 \end{aligned} \,. \qquad (11)$$

The coefficient S(N/2) is a sum of N/2 differences between two by two neighbouring points. In order to be used as an indication of local defects of objects with different size of the description this symptom must also be normalized relative to N:

$$S^*(N/2) = \frac{1}{N}\sum_{i=0}^{N-1} s(i)W_N^{(N/2).i} \,. \qquad (12)$$

For each of the coefficients of DFT can be made similar analyzes. The result can be summarized according to the proposed equation for determination the symptoms for recognition in accordance with the size of the description of each object:

$$S^*(n) = \frac{1}{N}\sum_{i=0}^{N-1} s(i)W_N^{n.i} \,. \qquad (13)$$

Using this approach, each object will contain different number of symptoms, based on the number of actual reported values in its description. In the synthesis of the classifiers in this peculiarity can be approached in two ways. The first way is to determine the minimum possible number of symptoms of all objects and to choose this number for the symptoms defined by equation (13).

The second way is to use adapted, to fit into a different number of symptoms, pattern recognition methods (PRM). For example, when adjusting the (PRM) "k nearest neighbours" (Fukunaga, Witten), the Euclidean distance as a measure of the distance of the unknown object A and a reference image s, can be determined:

Journal of Agricultural Studies
ISSN 2166-0379
2015, Vol. 3, No. 1

$$L(A,s)=\left\{\frac{1}{N_{min}}\sum_{i=0}^{N_{min}-1}\left(S_A^*(i)-S^*(i)\right)^2\right\}^{1/2}, \qquad (8)$$

where $N_{min}=\min(N_A, N)$, N_A – number of reported values of the description of A, N – number of reported values of the description of s.

The analysis of the Fourier transform, when used for synthesis of symptoms for recognition, can be spread similarly to other orthogonal transformations (Cosine, Sine, Hartley, Haar and others) that enable the execution of the algorithm in any number of reporting points of the primary description.

3. Results and Discussion

For experimental verification of performance (fitness) of the adapted PRM are used data [4, 6] for the values of light transmittance coefficient, obtained as the ratio of the signals from the photo-electronic sensors at two different wavelengths and λ_1 and λ_2 and longitudinal scanning of whole unpeeled potato tubers of variety "Agria", classified by experts in 3 classes (Georgiev). The maximum number of reported values in the descriptions is 25, which determines the maximum number of symptoms (given symmetry module coefficients definition in (13)) that can be used – in this case it is 13.

The available set of general studied objects contains 1145 objects of first class, 1165 - from the second and 880 in third. Randomly without re-election is formed a training set of 300 objects in class and from the remaining sets of the classes is formed a control set, also at random without re-election, containing 550 objects of each class.

Experiments were conducted with synthesized classifiers using as symptoms for recognition the coefficients defined in (13), and including different number of symptoms. Classifiers using the symptoms as follows: S*(0); S*(0) and S*(1); from S*(0) to S*(2), ..., from S*(0) to S*(12), were tested.

To evaluate the performance of classifiers is used the parameter "error of classification for a class" that evaluates incorrectly classified objects in the class. The two class classifier has two types' errors – error of higher quality class and error of lower quality class. Errors due to transition from the class with lower quality to that with better, the term of the theory of statistical decisions, can be classified as errors of second kind (omission of the target) and errors due to the transition from best to worst class – as the error of first kind (false alarm). In determining the errors was used approach to compare the classes against each other. In the case of three classes' classifier, the important classification mappings are defined–first class against second class, first class against third class, second class against the third class.

For presenting the results of the comparison of class i against class j (i/j) are used designations: $F_{i/j}^{I}$ – the error of first kind, $F_{i/j}^{II}$ – the error of the second kind.

The results for the errors of the first and second kind from tested classifiers, differing in the number of used symptoms for recognition, are presented in Table 1. In classifiers utilizing signs S*(0), S*(0) and S*(1), from S*(0) to S*(2), ..., from S*(0) to S*(4) classical method for pattern recognition "k nearest neighbours" is applied. The minimum number of reported values in the descriptions of the studied objects are 9 and therefore each will have at least symptoms from S*(0) to S*(4). And in the other cases, as the descriptions of the objects in the feature space have a different number of features, it is used the adapted method of pattern recognition "k nearest neighbours".

The results presented in Table 1 show that the errors of the first and second kind of classifiers of potato tubers slightly decrease with increasing the number of symptom. There is also a saturation– the values of the various types of errors for the cases using the symptoms $S^*(0) \div S^*(10)$, $S^*(0) \div S^*(11)$ and $S^*(0) \div S^*(12)$ are the same. Which means that for this case the symptoms S*(11) and S*(12) can be excluded.

Table 1. Errors of the first and second kind of classifiers using different number of scaled coefficients of DFT as symptoms for recognition

Symp-toms / Errors	$S^*(0)$	$S^*(0)$ and $S^*(1)$	$S^*(0) \div S^*(2)$	$S^*(0) \div S^*(3)$	$S^*(0) \div S^*(4)$	$S^*(0) \div S^*(5)$	$S^*(0) \div S^*(6)$
$F^{I}_{1/2}$	0,10103	0,10095	0,10000	0,10000	0,10000	0,10000	0,10000
$F^{I}_{1/3}$	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
$F^{I}_{2/3}$	0,13906	0,13701	0,13497	0,13469	0,13469	0,13469	0,13469
$F^{II}_{1/2}$	0,12656	0,12629	0,12603	0,12397	0,12397	0,12397	0,12397
$F^{II}_{1/3}$	0,00416	0,00417	0,00417	0,00415	0,00415	0,00415	0,00415
$F^{II}_{2/3}$	0,12791	0,12774	0,12774	0,12409	0,12409	0,12409	0,12409

Symp-toms / Errors	$S^*(0) \div S^*(7)$	$S^*(0) \div S^*(8)$	$S^*(0) \div S^*(9)$	$S^*(0) \div S^*(10)$	$S^*(0) \div S^*(11)$	$S^*(0) \div S^*(12)$
$F^{I}_{1/2}$	0,10000	0,10000	0,10000	0,09818	0,09818	0,09818
$F^{I}_{1/3}$	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
$F^{I}_{2/3}$	0,13469	0,13469	0,13265	0,13265	0,13265	0,13265
$F^{II}_{1/2}$	0,12397	0,12397	0,12371	0,12371	0,12371	0,12371
$F^{II}_{1/3}$	0,00415	0,00415	0,00415	0,00415	0,00415	0,00415
$F^{II}_{2/3}$	0,12409	0,12409	0,12409	0,12409	0,12409	0,12409

5. Conclusion

Agricultural products are organic objects with complex characteristics. Their special feature is that they have different geometric dimensions. This fact creates disadvantages for using certain methods for synthesis of symptoms and/or methods for pattern recognition. In the work is proposed to adapt the spectral method for the synthesis of symptoms using DFT and also to adapt the method of pattern recognition "k nearest neighbours". Classifiers were tested using the adapted methods with a different number of symptoms for recognition. The results confirm the statement that coefficients with larger numbers of DFT wear less energy of the signal and are therefore less informative as symptoms, and may be neglected.

References

Donevska, S., & Donevski, B. (1999). Fast Fourier Transform, Sofia, BG

Fukunaga, K. (1990). Introduction to Statistical Pattern Recognition. Academic Press Professional, Inc. San Diego, CA, USA

Georgiev, A., & Kostadinova, L. (1985). Automated systems for qualification and sorting agricultural products, Journal "Food Industry" №6, Sofia, BG

Georgiev, A., Kostadinova, L. & Gabrova, R. On-line Sorting Apples in Modern Technologies for Providing Quality of Fruit and Vegetable in Chains. Acta Horticulturae, N 712, Proc. of the IV Intern. *Conf. on Managing Quality in Chains, 2*, 911-914, MQUIC2006, Bangkok, Thailand

Witten, Ian, H., & Frank, Eibe. (2005). Data Mining Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA

Copyright Disclaimer